An Iteratively Reweighted Norm Algorithm for Minimization of Total Variation Functionals

Brendt Wohlberg* and Paul Rodríguez

EDICS: IMD-ANAL

***Abstract*— Total Variation (TV) regularization has become a very popular method for a wide variety of image restoration problems, including denoising and deconvolution. Recently, a number of authors have noted the advantages, including superior performance with certain non-Gaussian noise, of replacing the standard ℓ^2 data fidelity term with an ℓ^1 norm. We propose a simple but very flexible and computationally efficient method, the Iteratively Reweighted Norm algorithm for solving the generalized TV functional, which includes the ℓ^2-TV and and ℓ^1-TV problems.**

***Index Terms*— image restoration, inverse problem, regularization, total variation**

I. Introduction

Total Variation (TV) regularization has become a very popular method for a wide variety of image restoration problems, including denoising and deconvolution [1], [2]. The standard TV regularized solution of the inverse problem involving data $\mathbf{s}$ and forward linear operator K (the identity in the case of denoising, and a convolution for a deconvolution problem, for example) is the minimum of the functional

$$T(\mathbf{u}) = \frac{1}{2}\left\|K\mathbf{u} - \mathbf{s}\right\|_2^2 + \lambda \left\|\sqrt{(D_x\mathbf{u})^2 + (D_y\mathbf{u})^2}\right\|_1 , \quad (1)$$

where we employ the following notation:

- the p-norm of vector $\mathbf{u}$ is denoted by $\|\mathbf{u}\|_p$,
- scalar operations applied to a vector are considered to be applied element-wise, so that, for example, $\mathbf{u} = \mathbf{v}^2 \Rightarrow u_k = v_k^2$ and $\mathbf{u} = \sqrt{\mathbf{v}} \Rightarrow u_k = \sqrt{v_k}$, and
- horizontal and vertical discrete derivative operators are denoted by D_x and D_y respectively.

While a number of algorithms [3], [4] have been proposed to solve this optimization problem, it remains a computationally expensive task which can be prohibitively costly for large problems and non-sparse forward operator K.

Recently, the modified TV functional with an l^1 data fidelity term

$$T(\mathbf{u}) = \left\|K\mathbf{u} - \mathbf{s}\right\|_1 + \lambda \left\|\sqrt{(D_x\mathbf{u})^2 + (D_y\mathbf{u})^2}\right\|_1 . \quad (2)$$

has attracted attention [5], [6] due to a number of advantages, including superior performance with non-Gaussian noise such as salt and pepper noise, and applications in cDNA microarray image processing [7], illumination normalization [8], and cartoon-texture decomposition [9]. The standard approaches to solving problem (1) are not effective for problem (2), for which algorithm development is not well advanced [6].

We propose a simple but computationally efficient and very flexible method for solving the generalized TV functional

$$T(\mathbf{u}) = \frac{1}{p}\left\|K\mathbf{u} - \mathbf{s}\right\|_p^p + \frac{\lambda}{q} \left\|\sqrt{(D_x\mathbf{u})^2 + (D_y\mathbf{u})^2}\right\|_q^q \quad (3)$$

for $p \geq 1$ and $q \geq 1$, including both problems (1) and (2).

II. Iteratively Reweighted Norm Approach

Our Iteratively Reweighted Norm (IRN) approach is motivated by the Iteratively Reweighted Least Squares (IRLS) [10] algorithm for solving the minimum l^p norm problem

$$\min_{\mathbf{u}} \frac{1}{p}\left\|K\mathbf{u} - \mathbf{s}\right\|_p^p \quad (4)$$

by solving a sequence of minimum weighted l^2 norm problems, and is also closely related to the Iterative Weighted Norm Minimization algorithms [11], [12] for sparse signal decompositions.

These methods approximate the l^p norm of $\mathbf{u}$

$$\frac{1}{p}\left\|\mathbf{u}\right\|_p^p = \frac{1}{p}\sum_k |u_k|^p ,$$

by the weighted l^2 norm of $\mathbf{u}$

$$\frac{1}{2}\left\|W^{1/2}\mathbf{u}\right\|_2^2 = \frac{1}{2}\mathbf{u}^T W \mathbf{u} = \frac{1}{2}\sum_k w_k u_k^2$$

with diagonal weight matrix

$$W = (2/p)\operatorname{diag}\left(|\mathbf{u}|^{p-2}\right) . \quad (5)$$

To simplify somewhat, this approximation may be used to minimize the norm because, for the same choice of W (and $\mathbf{u}$ such that $u_k \neq 0\ \forall k$) we have

$$\nabla_{\mathbf{u}} \frac{1}{p}\left\|\mathbf{u}\right\|_p^p = (p/2)\nabla_{\mathbf{u}} \frac{1}{2}\left\|W^{1/2}\mathbf{u}\right\|_2^2 ,$$

so that both expressions have the same value and tangent direction. A detailed discussion of convergence issues for the IRLS problem may be found in [13], and we are currently preparing a paper including convergence details for our IRN method.

Brendt Wohlberg is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: brendt@t7.lanl.gov, Tel: (505) 667 6886, Fax: (505) 665 5757

Paul Rodríguez is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: prodrig@t7.lanl.gov, Tel: (505) 606 1483 , Fax: (505) 665 5757

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396 and was partially supported by the NNSA's Laboratory Directed Research and Development Program.

A. Data Fidelity Term

The data fidelity term of Equation (3) has the form of the IRLS functional in Equation (4), and is handled in the same way, representing

$$\frac{1}{p}\left\|K\mathbf{u}-\mathbf{s}\right\|_p^p \quad \text{by} \quad \frac{1}{2}\left\|W_F^{1/2}\left(K\mathbf{u}-\mathbf{s}\right)\right\|_2^2$$

with iteratively updated weights W_F. Since the choice defined by Equation (5) gives infinite weights for $p < 2$ and $u_k = 0$, we set

$$W_F = \operatorname{diag}\left(\frac{2}{p} f_F(K\mathbf{u}-\mathbf{s})\right)$$

where

$$f_F(x) = \begin{cases} |x|^{p-2} & \text{if } |x| > \epsilon_F \\ \epsilon_F^{p-2} & \text{if } |x| \leq \epsilon_F, \end{cases}$$

for some small ϵ_F, a common approach for IRLS algorithms [10].

B. Regularization Term

It is not quite as obvious how to express the TV regularization term from Equation (3) as a weighted l^2 norm. Given vectors $\mathbf{u}$ and $\mathbf{v}$ we have (using block-matrix notation)

$$\left\|\begin{pmatrix} W^{1/2} & 0 \\ 0 & W^{1/2} \end{pmatrix}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}\right\|_2^2 = \sum_k w_k u_k^2 + w_k v_k^2$$

so that when

$$W = \operatorname{diag}\left(\frac{2}{q}\left(\mathbf{u}^2+\mathbf{v}^2\right)^{(q-2)/2}\right)$$

we have

$$\frac{1}{2}\left\|\begin{pmatrix} W^{1/2} & 0 \\ 0 & W^{1/2} \end{pmatrix}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}\right\|_2^2 = \frac{1}{q}\left\|\sqrt{\mathbf{u}^2+\mathbf{v}^2}\right\|_q^q.$$

We therefore define the operator D and weights $\tilde{W}$

$$D = \begin{pmatrix} D_x \\ D_y \end{pmatrix} \qquad \tilde{W} = \begin{pmatrix} W & 0 \\ 0 & W \end{pmatrix}$$

so that $\|\tilde{W_R}^{1/2} D\mathbf{u}\|_2^2 = \|W_R^{1/2} D_x\mathbf{u}\|_2^2 + \|W_R^{1/2} D_y\mathbf{u}\|_2^2$ with weights defined by

$$W_R = \operatorname{diag}\left(\frac{2}{q}\left((D_x\mathbf{u})^2+(D_y\mathbf{u})^2\right)^{(q-2)/2}\right)$$

gives the desired term. (Note that this is *not* the separable approximation, as in [14], for example, which is often used.)

As in the case of the data fidelity term, care needs to be taken when $q < 2$ and $u_k = 0$. We define

$$f_R(x) = \begin{cases} |x|^{(q-2)/2} & \text{if } |x| > \epsilon_R \\ 0 & \text{if } |x| \leq \epsilon_R, \end{cases}$$

for some small ϵ_R, and set

$$W_R = \operatorname{diag}\left(\frac{2}{q} f_R\left((D_x\mathbf{u})^2+(D_y\mathbf{v})^2\right)\right).$$

Note that f_R sets values smaller than the threshold, ϵ_R, to zero, as opposed to f_F, which sets values smaller than the threshold, ϵ_F, to ϵ_F^{p-2}. Our motivation for this choice is that a region with very small or zero gradient should be allowed to have zero contribution to the regularization term, rather than be clamped to some minimum value. In practice, however, we have found that this choice does not give significantly different results than the standard IRLS approach represented by f_F.

C. General Algorithm

Combining the terms described in Sections II-A and II-B, we have the functional

$$T(\mathbf{u}) = \frac{1}{2}\left\|W_F^{1/2}\left(K\mathbf{u}-\mathbf{s}\right)\right\|_2^2 + \frac{\lambda}{2}\left\|\tilde{W_R}^{1/2} D\mathbf{u}\right\|_2^2$$

which, it is worth noting, may be expressed as

$$T(\mathbf{u}) = \frac{1}{2}\left\|\begin{pmatrix} W_F^{1/2} & 0 \\ 0 & \tilde{W_R}^{1/2} \end{pmatrix}\left(\begin{pmatrix} K \\ \sqrt{\lambda} D \end{pmatrix}\mathbf{u} - \begin{pmatrix} \mathbf{s} \\ 0 \end{pmatrix}\right)\right\|,$$

which has the same form as an IRLS problem, but differs in the computation of $\tilde{W_R}^{1/2}$. The minimum of this functional is

$$\mathbf{u} = \left(K^T W_F K + \lambda D^T \tilde{W}_R D\right)^{-1} K^T W_F^{1/2}\mathbf{s}, \qquad (6)$$

and the resulting algorithm consists of the following steps:

Initialize

$$\mathbf{u}_0 = \left(K^T K + \lambda D^T D\right)^{-1} K^T \mathbf{s}$$

Iterate

$$\begin{aligned} W_{F,k} &= \operatorname{diag}\left(\frac{2}{p} f_F(K\mathbf{u}_{k-1}-\mathbf{s})\right) \\ W_{R,k} &= \operatorname{diag}\left(\frac{2}{q} f_R\left((D_x\mathbf{u}_{k-1})^2+(D_y\mathbf{u}_{k-1})^2\right)\right) \\ \mathbf{u}_k &= \left(K^T W_{F,k} K + \lambda D_x^T W_{R,k} D_x \right. \\ &\quad \left. + \lambda D_y^T W_{R,k} D_y\right)^{-1} K^T W_{F,k}^{1/2}\mathbf{s} \end{aligned}$$

The matrix inversion is achieved using the Conjugate Gradient (CG) method. We have found that a significant speed improvement may be achieved by starting with a high CG tolerance which is decreased with each main iteration until the final desired value is reached. (We are currently preparing a paper which addresses convergence issues in detail.)

III. l^1 Data Fidelity

In the remainder of this paper we shall restrict our attention to the l^1-TV case ($p = 1$, $q = 1$), but note that this flexible approach is capable of efficiently solving other cases as well, including the standard l^2-TV case ($p = 2$, $q = 1$) where we have found it to be slightly slower than the lagged diffusivity algorithm [3], to which it is related.

For the results reported here, operators D_x and D_y were defined by applying the same one-dimensional discrete derivative along image rows and columns respectively. Applied to vector $\mathbf{u} \in \mathbb{R}^N$, this discrete derivative was computed as $u_k - u_{k+1}$ for the derivative at index $k \in \{0, 1, \ldots, N-2\}$, and as $(u_{N-3} - 8u_{N-2})/12$ for the derivative at index $N-1$. Values in the ranges 10^{-2} to 10^{-4} and 10^{-4} to 10^{-8} were used for constants ϵ_F and ϵ_R respectively. All program run times were obtained on a 2.8GHz Intel Xeon processor.

A. Denoising

We first consider the denoising problem corresponding to the choice $K = I$. A comparison of l^2 TV (computed using the lagged diffusivity algorithm) and l^1 TV (computed using the IRN method) denoising performance for speckle noise is displayed in Figure 1. (The original image in Figure 1(a) was scaled by a factor of 1/255 to give a nominal pixel value range of 0 to 1.) Note the very significantly superior performance of l^1 TV denoising for this type of noise. In the l^2 TV case, edges are better preserved when using a smaller weighting, λ, to the regularization term, but at the expense of noise reduction and SNR.

Direct application of equation (6) for l^1 TV denoising is significantly slower than l^2 TV via lagged diffusivity as a consequence of the significantly greater number of CG iterations required to solved the system for each main iteration. When $K = I$, however, we may apply the substitution $\tilde{\mathbf{u}} = W_F^{1/2}\mathbf{u}$, giving

$$T(\tilde{\mathbf{u}}) = \frac{1}{2}\left\|\tilde{\mathbf{u}} - W_F^{1/2}\mathbf{s}\right\|_2^2 + \frac{\lambda}{2}\left\|\tilde{W_R}^{1/2} D W_F^{-1/2}\tilde{\mathbf{u}}\right\|_2^2,$$

with solution

$$\tilde{\mathbf{u}} = \left(I + \lambda W_F^{-1/2} D^T \tilde{W}_R D W_F^{-1/2}\right)^{-1} W_F^{1/2}\mathbf{s}.$$

Applying this modification to the general algorithm results in a significant reduction in the required number of CG iterations, a comparison of which is provided in Figure 2. The run times for the lagged diffusivity, and the direct and indirect IRN algorithms were 4.2s, 41.1s, and 10.8s respectively.

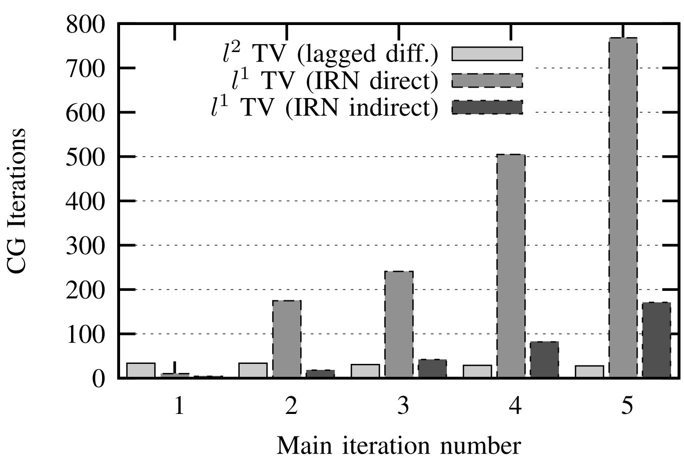

Fig. 2. A comparison of CG iterations for lagged diffusivity, and the direct and indirect IRN algorithms for denoising.

B. Deconvolution

We apply the IRN algorithm to the problem of deconvolution of an image convolved by a separable smoothing filter having 9 taps and approximating a Gaussian with standard deviation of 2.0. In this case K is the corresponding linear operator, and the substitution applied in the previous section is no longer possible. We constructed a test image by convolving the image in Figure 1(a) by the smoothing kernel, and adding 5% speckle noise, giving an image with an SNR of 3.4dB. Comparing the performance of l^2 TV (computed using the lagged diffusivity algorithm) and l^1 TV (computed using the IRN method) deconvolution, we obtain a 12.2dB reconstruction in 14.2s and a 14.4dB reconstruction in 38.6s respectively. A comparison of the numbers of CG iterations required by each method is provided in Figure 3.

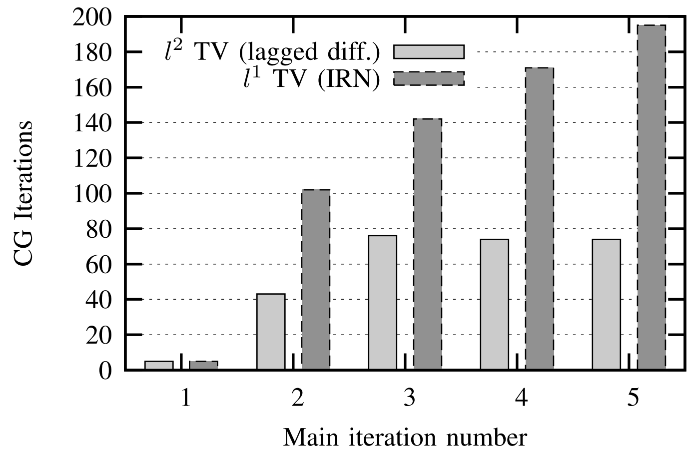

Fig. 3. A comparison of CG iterations for lagged diffusivity and the IRN algorithms for deconvolution.

IV. Conclusions

The Iterative Reweighted Norm (IRN) approach provides a simple but computationally efficient method for TV regularized optimization problems, including both denoising and those such as deconvolution having a linear operator in the data fidelity term. This framework is very flexible, and can be applied to regularized inversions with a wide variety of norms for the data fidelity and regularization terms, including the standard l^2 TV, and more recently proposed l^1 TV formulations. This method provides a significantly faster algorithm for the l^1 TV formulation than any other algorithm of which we are aware. We are currently working on developing preconditioning strategies to further improve the speed by reducing the number of CG iterations required, and have already obtained a factor of two improvement over the best denoising run time reported in Section III-A by using a Jacobi line relaxation preconditioner [15, Chapter 7, pp. 123]. A software implementation [16] is available under an open-source license.

Acknowledgment

The authors thank Markus Berndt for valuable discussion on CG preconditioning methods.

References

[1] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, June 1998.

[2] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *The Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras., Eds. Springer, 2005.

(a) Original 512×512 image

(b) Image with 5% speckle noise. SNR: 3.9dB.

(c) Denoised using l^2 TV ($\lambda = 0.20$). SNR: 11.9dB.

(d) Denoised using l^1 TV ($\lambda = 1.3$). SNR: 20.3dB.

Fig. 1. A comparison of l^2 TV and l^1 TV denoising for speckle noise. The l^2 TV results were computed using the lagged diffusivity algorithm [3] while the l^1 TV results were computed using the IRN algorithm.

[3] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1-4, pp. 227–238, Jan. 1996.

[4] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[5] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.

[6] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L^1 function approximation," *SIAM Journal on Applied Mathematics*, vol. 65, no. 5, pp. 1817–1837, 2005.

[7] W. Yin, T. Chen, X. S. Zhou, and A. Chakraborty, "Background correction for cDNA microarray images using the TV+L^1 model," *Bioinformatics*, vol. 21, no. 10, pp. 2410–2416, 2005.

[8] T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. Huang, "Illumination normalization for face recognition and uneven background correction using total variation based image models," in *Proceedings of IEEE confererence on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 532–539.

[9] W. Yin, D. Goldfarb, and S. Osher, "Image cartoon-texture decomposition and feature selection using the total variation regularized L^1 functional," *Lecture Notes in Computer Science*, vol. 3752, pp. 73–84, 2005.

[10] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, no. 4, pp. 1071–1091, Oct. 1988.

[11] I. F. Gorodnitsky and B. D. Rao, "A new iterative weighted norm minimization algorithm and its applications," in *IEEE Sixth SP Workshop on Statistical Signal and Array Processing, 1992*, Victoria, BC, Canada, Oct. 1992.

[12] ——, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm." *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[13] K. P. Bube and R. T. Langan, "Hybrid ℓ^1/ℓ^2 minimization with applications to tomography," *Geophysics*, vol. 62, no. 4, pp. 1183–1195, July-August 1997.

[14] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 987–995, June 1996.

[15] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd ed. SIAM Books, 2000.

[16] P. Rodríguez and B. Wohlberg, "Numerical methods for inverse problems and adaptive decomposition (NUMIPAD)," Software library available from http://numipad.sourceforge.net/.